2014

Emerging of academic information search system with ontology-based approach

Norasykin Mohd Zaid
Universiti Teknologi Malaysia, nmz056@uowmail.edu.au

Sim Kim Lau
University of Wollongong, simlau@uow.edu.au

Publication Details

Mohd Zaid, N. & Lau, S. Kim. (2014). Emerging of academic information search system with ontology-based approach. In J. G. Laborda, F. Ozdamli & Y. Maasoglu (Eds.), 5th World Conference on Educational Sciences - WCES-2013 (pp. 132-138). Netherlands: Elsevier BV.

Emerging of academic information search system with ontology-based approach

Abstract
This paper describes the comparison of ontology development tools for development of academic information search system that assists inexperienced research students at a local university in Malaysia to search for academic resources in the local language context (Bahasa Malaysia). The cohort of inexperienced research students faces two main problems when using current system comprises of keyword search. Firstly the language barrier-limiting students' capabilities to conduct keyword search in foreign language (such as English). Secondly limited research experience in querying often results in obtaining irrelevant search results. The proposed semantic search system aims to apply ontology-based search to overcome the above two problems. The paper presents the first phase of system development; ontology design and ontology development tool.

Keywords
search, information, academic, emerging, approach, ontology, system

Publication Details
Mohd Zaid, N. & Lau, S. Kim. (2014). Emerging of academic information search system with ontology-based approach. In J. G. Laborda, F. Ozdamli & Y. Maasoglu (Eds.), 5th World Conference on Educational Sciences - WCES-2013 (pp. 132-138). Netherlands: Elsevier BV.

Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 116 (2014) 132 – 138

Procedia
Social and Behavioral Sciences

5th World Conference Educational Sciences -WCES 2013

Emerging of Academic Information Search System with Ontology-Based Approach

Norasykin Mohd Zaid [a 1], Sim Kim Lau [b]

[a]Faculty of Education, Universiti Teknologi Malaysia, Skudai 81310, Malaysia
[b]School of Information Systems and Technology, Faculty of Informatics, University of Wollongong, Wollongong 2500, Australia

Abstract

This paper describes the comparison of ontology development tools for development of academic information search system that assists inexperienced research students at a local university in Malaysia to search for academic resources in the local language context (Bahasa Malaysia). The cohort of inexperienced research students faces two main problems when using current system comprises of keyword search. Firstly the language barrier-limiting students' capabilities to conduct keyword search in foreign language (such as English). Secondly limited research experience in querying often results in obtaining irrelevant search results. The proposed semantic search system aims to apply ontology-based search to overcome the above two problems. The paper presents the first phase of system development; ontology design and ontology development tool.

Selection and/or peer-review under responsibility of Academic World Education and Research Center.
Keywords:

1. Introduction

Information and knowledge are increasingly becoming shareable and searchable resources, particularly in the current digitized world. Since 1996, the World Wide Web (WWW) has become a primary source for information offering online resources that are available 24/7. Traditionally library is an important source of information, particularly as academic resources and has become important source of reference for academic researchers. Library classification system has migrated from Dewey Decimal Classification System (DDC) to a new digitized format such as Online Public Access Catalog (OPAC) system that can be accessed through the web. The OPAC system is based on known-item search (Antelman et al., 2006).

Furthermore, keyword search and Boolean operators can also be used to facilitate the search process. Undoubtedly digital library provides an improved source of information access that include digital document creation and storage, documents classification and data indexing, documents searching and retrieving, distribution, administration and access control (Garza-Salazar et al., 2003). However human interpretation is still required when records matching the search criteria (such as keywords) are returned to determine its relevance and usefulness. For example, in searching for a programming textbook, which we do not know the exact title, we tend to type the word programming in the search box. When search results are returned, we scroll down the list of titles to look for the one that we search for. This is commonly encountered by students who are inexperienced in literature search. The motivation of this paper is to propose the development of an ontology-based information retrieval system to assist

[1]Corresponding Author: Norasykin Mohd Zaid. Tel.: +60-10-406-1060
E-mail address: norasykin@utm.my

Selection and/or peer-review under responsibility of Academic World Education and Research Center.
doi:10.1016/j.sbspro.2014.01.180

inexperienced research students at a local university in Malaysia to search for academic resources in the local language context (Bahasa Malaysia). The rest of the paper is organized as follows. Section 2 discusses ontology design and development of ontology. Section 3 discusses ontology development tools and follows by conclusion in Section 4.

2. Ontology

Ontologies have been known as a database of terms that justified a domain to be used and shared in a global area (Borst, 1997). Ontology becomes a model of real word to represent a domain of knowledge. This new technology has been used in the Semantic Web although the original word of ontology is being borrowed from the philosophy discipline, which defines the concepts of things. Thomas (1993) explains the real definition of ontology is a systematic account of existence, however in computer science, ontology is a representation of precise specification to form a concept. Thus, ontology is described as formal specification of terms in the define domain and identifying any relations existing in between the terms. Ontology enables people or machines to retrieve the desired information with an understanding of the meaning of terms used in the domain and share common vocabularies used in the same domain (Wang et al., 2008a). Therefore, the use of ontology is about using, reusing and sharing domain knowledge of terms concept. Many ontology classes have been developed recently and are kept in a database to be used or referred to by others as knowledge/resource sources. Ontologies are not only used in the field of Semantic Web but also in many others fields such as artificial intelligence, software engineering, biomedical informatics, library science, and information architecture.

There are two types of ontologies according to two dimensions of perception: the amount and type of structure of the conceptualisation and the subject of the conceptualisation. The first dimension, according to Heijst et al. (1995), includes: (i) terminological ontologies, (ii) information ontologies, and (iii) knowledge modelling ontologies; whereas the second dimension includes: (i) domain ontologies, (ii) generic ontologies, (iii) representation ontologies, and (iv) application ontologies. The first dimension with terminological ontologies is referred to as ontology that defines the terms to represent knowledge in the domain of discourse, such as medical or biological domains. Information ontologies are defined as records structure of a database, which is a flat structure, unlike the knowledge modeling ontologies, which have a richer structure of database, such as involving distinction and decision-making processes. To refer to the second dimension of ontologies, domain ontologies refer to specific particular area while generic ontologies refer to domain ontologies across many areas. Representation ontologies are supposed to be naturally present in general contrast to application ontologies, which are specifically designed to the particular application such as the Marine Metadata Interoperability Project (MMI) (https://marinemetadata.org/).

2.1. Ontology design

Holsapple and Joshi (2002) present five approaches to ontological design: (1) inspiration, (2) induction, (3) deduction, (4) synthesis, and (5) collaboration. Inspirational approach starts the design idea by collecting individual personal views and creativity to construct the domain context. Inductive approach is based on the observation and analysing of current or specific domains to apply to particular domains. Deductive approach adopts some general principles to construct a new domain while the synthetic approach applies some potential characterisation from the existing ontologies. With the collaborative approach, the approach relies on human participation, which involves individual reflection and viewpoints to get along with the collaborative process.

2.2. Ontology development

How these ontologies can be developed depends on how or what method is being used. Uschold and Gruninger (1996) conclude that there are five steps in the process of ontologies development: (i) identify purpose and scope, (ii) building the ontology, (iii) evaluation, (iv) documentation, and (v) guidelines for each phase. In the second step of building ontology, it includes: (a) ontology capture, (b) ontology coding, and (c) integrating existing ontologies (Uschold and Gruninger, 1996). The first step in building the ontology is by considering when there is a clear idea

on what ontology is going to build, and then the domain of the ontology can be set with purpose and scope of the domain identified earlier. This idea can then be extended to the second step of developing domain ontology by providing information of ontology capture, coding and with attention to consider using an existing ontology. The third step is important to identify whether the ontology is in a good form of classification and relationship in its domain to bring effectiveness of knowledge sharing. In the forth step, the idea of having documentation is to allow knowledge sharing by preparing the problems faced in existing ontology with the important assumption together with the concepts definition based on type and ontology purpose. In the last step, the initial guidelines are provided which consists of clarity, coherence and extensibility.

Some other methodologies for building ontology have also been discussed by Fernandez-Lopez et al. (1997); and Corcho et al. (2003a). Corcho et al. (2003a) have review and compare the main methodologies for building ontology such as METHONTOLOGY (Fernandez-Lopez et al., 1997) and On-To-Knowledge methodology (Steffen et al., 2001). Fernandez-Lopez et al. (1997) propose the ontology development process to start with planning, specifying, knowledge acquisition, conceptualising, formalising, integrating, implementing, evaluating, documenting and maintaining the process. This methodology is used in most ontology development processes (Lopez et al., 1999; and Brusa et al., 2008) and has also been extended to allow collaborative edition of ontologies at the knowledge level (Arpírez et al., 2001). On-To-Knowledge methodology takes into consideration the process of ontology development from the early stage of setting up the project until the final level of the application which consists of: feasibility study, ontology kickoff, refinement, evaluation and maintenance (Steffen et al., 2001).

Other than ontology methodology, there are currently many tools available to develop ontology. Some of them are: OilEd (Bechhofer et al., 2001), OntoEdit (Sure et al., 2002), WebODE (Corcho et al., 2003b), WebOnto (Domingue, 1998), Protégé (Rubin et al., 2007; Tudorache et al., 2008; Corcho et al., 2003b; and Corcho and Gómez-Pérez, 2004), OntoSaurus (Swartout et al., 1997), Ontolingua (Thomas, 1993; and Farquhar et al., 1997b), KAON (Bozsak et al., 2002) and SymOntoX (Missikoff and Taglino, 2003). For example, OntoSaurus (Swartout et al., 1997) presents ontology browsers and editors to support a collaborative vision of ontology development. Farquhar et al. (1997a) develop tools to allow ontology sharing and at the same time provide services to publish, browse, create and ontologies editing stored on their Ontolingua server.

Table 1 Ontology Development Tools

	Ontology Development Tools	Main Features	Ontology Development Process	Strength	Weakness
1.	OilEd	Ontology editor for building ontologies using Ontology Interchange Language (OIL).	IOilEd knowledge model is based on description logics.	Allow users to exploit the full power of an expressive web ontology language (OIL/DAML+OIL). Reasoning is used to support the design and maintenance of ontologies.	No support for versioning or for working with multiple ontologies. The reasoning support provided (FaCT) is incomplete for OIL extended with concrete data-types and individuals, and does not include additional services such as explanation.
2.	OntoEdit	Ontology editor	Methodology-guided development of ontologies.	Ontology development with the help of inferencing. Extensibility through plug-in structure.	No support on built-in inference engine, DBMS, collaborative working and ontology library.
3.	WebODE	Integrated technological support for many activities of the ontology lifecycle, technological support for ontology	Provides a default form-based web user interface to create ontologies according to the knowledge model aforementioned.	Extensibility via plug-in. Graph view. Multi-user. Merging.	No support on ontology library.

		development methodologies, ontology interoperability.			
4.	WebOnto	A tool for collaboratively browsing and editing ontologies, which uses a Java client to alleviate the interface problems generated by HTML.	Provides direct manipulation interface using graphical representations to present ontology constructs.	Incorporates the sketching and synchronous communication tools. Provides a direct manipulation interface using graphical representations to present ontology constructs.	No support on extensibility and merging.
5.	Protégé	An open-source tool that allows developers to create and to manage terminologies and ontologies.	Protégé knowledge model is frame-based.	Protégé 2000 provides ontology editing functionality on different levels. It is possible to query the ontology. Plug-ins are provided for querying based on F-Logic, merging and annotation of the ontologies with WordNet.	Does not provide real support for multiple users. There is no support for multiple changes on the same component. Users are not notified about changes made by others. The tool does not allow two classes or attributes with the same name.
6.	OntoSaurus	Ontology browser.	Linking domain specific terms to an existing ontology and extending it.	Share knowledge across systems; use object-oriented Common-Lisp-based as a web server and data knowledge can be browsed and edited on any platform.	Use HTML for the client interface that generates interface design problems. It does not support ontology development from scratch.
7.	Ontolingua	Provides a suite of ontology authoring tools and a library of modular re-useable ontologies.		Enables renaming of non-logical symbols from multiple component ontologies and that disambiguate symbol references during input and output.	No support for graph view and extensibility.
8.	KOAN	Builds on available resources and provides tools for the engineering, discovery, management, and presentation of ontologies and metadata.	Includes a multi-lingual Ontology Engineering and Evolution Environment (OntoMatSOEP) that allows the manual development and maintenance of ontologies.	Provides objects representing various pieces of ontology, such as Concept, Relation, Instance, Objects for creating and applying changes to ontology entities as well as objects providing query facilities.	No support on graph view and merging.
9.	SymOntoX	Ontology management system that is capable of developing and managing several ontologies (business and enterprise).		Support collaborative and distributed ontology authoring activities.	
10.	Retrievalware 8.0	Is an enterprise search engine emphasising natural language processing and semantic networks.	Combines Convera's proven enterprise search and categorisation capabilities with a new dynamic	Uses a semantic network to expand queries for more complete recall. Profiling feature filters content to enable real-	

	classification methodology to help organisations automate knowledge management and discovery processes.	time monitoring of information in live data sources.

3. Ontology design and development tool

The ontology-based search system is developed based on an ontology-based mind-map. The mind-map is developed from the academic programme profile of the faculty in this case study. The Education Faculty aims to produce future teachers with knowledge and experiences related to the teaching profession. Thus, research topics are often conducted on issues related to teaching and learning based on specialised and professional subjects offered at the faculty. The mind-map is developed by considering the relationship between major components of teaching and learning as specialised subject offering, for example mathematics, physics, chemistry, living skills, sports science, Islamic studies, computer science and Teaching Language as a Second Language (TESL).

The mind-map is organised in a hierarchical structure to be translated to an ontology structure. The mind-map is developed using the inductive approach of ontology design (Holsapple and Joshi, 2002). With an inductive approach, the researcher observes, examines and analyses the sample domain of interest to develop the required ontology.

To prepare for mind-map development, the author examines the research and teaching areas in the faculty. Five main categories of teaching and learning that match the faculty profile are identified: teaching, learning, field of study, education level and others. For each main category, subcategories are prepared to show relevant topic areas in each category. For example, the teaching main category consists of the following subcategories: pedagogy, educator, skill, style, course or subject, theory and tool. To illustrate the subcategory of pedagogy relates to research conducted on a thesis, which investigates how the teaching pedagogy is applied. The same rationale applies for the learning category. In the case of field of study category, it reflects the course or subject offered in the faculty. These include mathematics, physics, chemistry, living skills, sports science, Islamic studies, computer science and TESL. The same category is also included as subcategories of teaching and learning. These examples of cross-categories options enable users to select the same keyword, but under a different structure of category-subcategory, to give a variety of combinations to enable users to choose from different perspectives.

As explained, the proposed ontology development is based on inductive approach, which is based on the observation and analysing of current or specific domains to apply to particular domains. The prototype development is based on existing library records in the database system of Faculty of Education, Universiti Teknologi Malaysia. The thesis records obtained for the prototype development was derived from the database system in 2008. Thus using the inductive approach the subcategories created in the prototype system are based on existing structure of the database system. It is worth noting that the subcategories structure can be changed whenever a new structure of categories and subcategories is needed or requires modification.

Another unique characteristic found in this university is that the thesis can be written in the native language of Bahasa Malaysia or the English language. As a non-English speaking country, this university conducts most of the courses in the Bahasa Malaysia language. However, courses can be taught in the English language, particularly for students who are enrolled in English teaching courses such as TESL or international student. These students write their thesis in English; therefore it is desirable for the search system to be able to conduct searches for thesis records regardless of language.

Protégé 3.4.4 is used to develop the ontology classes and subclasses. Protégé is a tool to help users in the construction of a small to large knowledge repository. Protégé represents the concepts and relationships as mind-map or concept map features, using an easy to use graphical representation and it allows users to have direct manipulation such as content controlling and editing. This way, the knowledge browsing and editing process can become simple and flexible.

4. Conclusion

Ontology enables relationships between keywords and terms to be defined. Ontology allows desired information to be retrieved by sharing common vocabularies with an understanding of meaning of terms in the domain. The ontology design is based on an ontology-based mind-map, which reflect to the real database system of selected domain. Protégé 3.4.4 is used as the ontology development tool to develop the ontology classes and subclasses based on the designed mind-map.

Acknowledgements

The authors would like to thank Universiti Teknologi Malaysia and University of Wollongong for their support in making this research possible. This research was supported by Malaysian government (Ministry of Higher Education) and Universiti Teknologi Malaysia (UTM).

References

Antelman, K., Lynema, E. and Pace, A. K. (2006). Toward a Twenty-First Century Library Catalog *Information Technology & Libraries* 25 (3), 128-139.

Arpírez, J. C., Corcho, O., Fernández-López, M. and Gómez-Pérez, A. (2001). Webode: A Scalable Workbench for Ontological Engineering. *Proceeding of the 1st International Conference on Knowledge Capture*, Victoria, British Columbia, Canada**,** 6-13.

Bechhofer, S., Horrocks, I., Goble, C. and Stevens, R. (2001). Oiled: A Reason-Able Ontology Editor for the Semantic Web. *Ki 2001: Advances in Artificial Intelligence.* Springer Berlin/Heidelberg.

Borst, W. N. (1997). Construction of Engineeering Ontologies for Knowledge Sharing and Reuse. University of Twente.

Bozsak, E., Ehrig, M., Handschuh, S., Hotho, A., Maedche, A., Motik, B., Oberle, D., Schmitz, C., Staab, S., Stojanovic, L., Stojanovic, N., Studer, R., Stumme, G., Sure, Y., Tane, J., Volz, R. and Zacharias, V. (2002). Kaon - Towards a Large Scale Semantic Web. *Third International Conference on E-Commerce and Web Technologies*, London, UK**,** 304 - 313.

Brusa, G., Laura Caliusco, M. and Chiotti, O. (2008). Towards Ontological Engineering: A Process for Building a Domain Ontology from Scratch in Public Administration. *Expert Systems,* 25 (5), 484-503. DOI:10.1111/j.1468-0394.2008.00471.x.

Corcho, O., Fernandez-Lopez, M. and Gomez-Perez, A. (2003a). Methodologies, Tools and Languages for Building Ontologies: Where Is Their Meeting Point? *Data & Knowledge Engineering,* 46(1), 41-64.

Corcho, O. and Gómez-Pérez, A. (2004). Ontology Translation Approaches for Interoperability: A Case Study with Protégé-2000 and Webode. *The 14th International Conference on Knowledge Engineering and Knowledge Management - EKAW'04*, Northamptonshire, UK**,** 16.

Corcho, Ó., Gómez-Pérez, A., Guerrero-Rodríguez, D. J., Pérez-Rey, D., Ruiz-Cristina, A., Sastre-Toral, T. and Suárez-Figueroa, M. C. (2003b). Evaluation Experiment of Ontology Tools' Interoperability with the Webode Ontology Engineering Workbench. *The 2nd International Semantic Web Conference (ISWC'03)*, Sanibel Island, Florida, USA, 87.

Domingue, J. (1998). Tadzebao and Webonto: Discussing, Browsing, and Editing Ontologies on the Web. *11th Knowledge Acquisition for Knowledge-Based Systems Workshop*.

Farquhar, A., Fikes, R. and Rice, J. (1997a). The Ontolingua Server: A Tool for Collaborative Ontology Construction. *International Journal of Human-Computer Studies,* 46 (6), 20.

Farquhar, A., Fikes, R. and Rice, J. (1997b). Tools for Assembling Modular Ontologies in Ontolingua. *The Fourteenth National Conference on Artificial Intelligence (AAAI'97)*, Providence, Rhode Island.

Fernandez-Lopez, M., Gomez-Perez, A. and Juristo, N. (1997). Methontology: From Ontological Art Towards Ontological Engineering. *Proceedings of the AAAI97 Spring Symposium* Stanford, USA.

Garza-Salazar, D. A., Lavariega, J. C. and Sordia-Salinas, M. (2003). Information Retrieval and Administration of Distributed Documents in Internet. IN Abramowicz, W. (Ed.) *Knowledge-Based Information Retrieval and Filtering from the Web.* USA, Kluwer Academic Publishers.

Heijst, G. V., Falasconi, S., Abu-Hanna, A., Schreiber, G. and Stefanelli, M. (1995). A Case Study in Ontology Library Construction. *Artificial Intelligence in Medicine,* 7 39.

Holsapple, C. W. and Joshi, K. D. (2002). A Collaborative Approach to Ontology Design. *Commun. ACM,* 45 (2), 42-47. DOI:10.1145/503124.503147.

Lopez, M. F., Gomez-Perez, A., Sierra, J. P. and Sierra, A. P. (1999). Building a Chemical Ontology Using Methontology and the Ontology Design Environment. *Intelligent Systems and their Applications, IEEE,* 14 (1), 37-46.

Missikoff, M. and Taglino, F. (2003). Symontox: A Web-Ontology Tool for Ebusiness Domain. *Proceedings of the 4th International Conference on Web Information Systems Engineering (WISE2003)*, Roma, Italy.

Rubin, D. L., Noy, N. F. and Musen, M. A. (2007). Protégé: A Tool for Managing and Using Terminology in Radiology Applications. *Journal of Digital Imaging,* 20, 12.

Steffen, S., Rudi, S., Hans-Peter, S. and York, S. (2001). Knowledge Processes and Ontologies. *IEEE Intelligent Systems,* 16 (1), 26-34. DOI:10.1109/5254.912382.

Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R. and Wenke, D. (2002). Ontoedit: Collaborative Ontology Development for the Semantic Web. *International Semantic Web Conference - ISWC'02*, Sardinia, Italy**,** 221–235.

Swartout, B., Patil, R., Knight, K. and Russ, T. (1997). Toward Distributed Use of Large-Scale Ontologies. *Symposium on Ontological Engineering.* USA, Association for the Advancement of Artificial Intelligence - AAAI.

Thomas, R. G. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition,* 5 (2), 199-220. DOI:http://dx.doi.org/10.1006/knac.1993.1008.

Tudorache, T., Noy, N. F. and Musen, M. A. (2008). Collaborative Protégé: Enabling Community-Based Authoring of Ontologies. *7th International Semantic Web Conference - ISWC'08*, Karlsruhe, Germany.

Uschold, M. and Gruninger, M. (1996). Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review.*

Wang, X., Fang, F. and Fan, L. (2008). Ontology-Based Description of Learning Object. *The 7th International Conference on Web-based Learning (ICWL 2008)*, Zhejiang Normal University, Jinhua, China**,** 468 - 476.